

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **10254496 A**

(43) Date of publication of application: **25 . 09 . 98**

(51) Int. Cl. **G10L 5/06**
G06F 15/18
G10L 3/00

(21) Application number: **09056162**

(71) Applicant: **MITSUBISHI ELECTRIC CORP**

(22) Date of filing: **11 . 03 . 97**

(72) Inventor: **ABE YOSHIHARU**

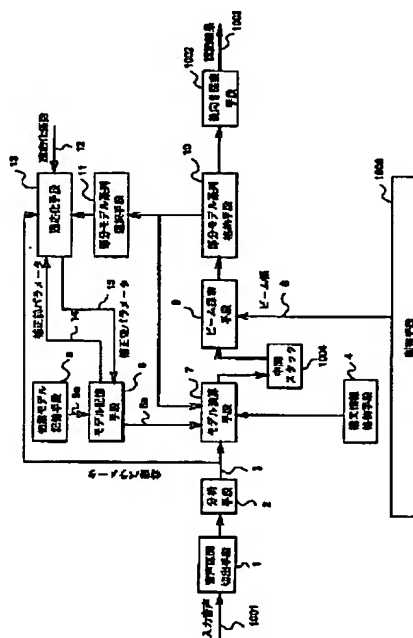
(54) **SPEECH RECOGNITION METHOD**

(57) Abstract:

PROBLEM TO BE SOLVED: To make it possible to reduce a search volume and also improve accuracy of speech recognition by determining a likelihood of an input speech to model group composed of adapted model groups by a search processing and configuring a system to recognize the speech based on this likelihood.

SOLUTION: A model calculating means 7 receives and input for one frame, and applies a parameter of a model in accordance with syntax information for each frame to the input and calculates the likelihood of the inputted feature parameter 3. Next, a partial model group selection means 11 selects a partial model group as information used for an adaptation at an adapting means 13. Next, after adaptation has been completed for all the selected models, the adapting means 14 adopts the parameters obtained from the result of adaptation as corrected parameters and replaces the parameters of a storage means 6 with the corrected parameters 15. And backward searching is operated by a backward searching means 1002 and a recognition result 1003 is obtained.

COPYRIGHT: (C)1998,JPO



【特許請求の範囲】

【請求項1】 入力音声の各フレームで入力音声にモデルを適応化し、この適応化されたモデルの系列からなるモデル系列に対する入力音声の尤度を探索処理により求め、この尤度に基づき音声認識を行うことを特徴とする音声認識方式。

【請求項2】 上記探索処理としてビーム探索を用いることを特徴とする請求項1記載の音声認識方式。

【請求項3】 入力音声の途中のビーム内の仮説の尤度の大きい方からN個の仮説を用いて、上記モデルを入力音声へ適応化することを特徴とする請求項2記載の音声認識方式。

【請求項4】 入力音声の途中のビーム内の仮説の尤度の大きい方からN個の仮説を用いるとともに、尤度に応じた重みによる加重を用いて上記モデルを入力音声へ適応化することを特徴とする請求項2記載の音声認識方式。

【請求項5】 上記モデルを音韻のモデルとしたことを特徴とする請求項1ないし4いずれかに記載の音声認識方式。

【請求項6】 上記モデルを音韻境界のモデルとしたことを特徴とする請求項1ないし4いずれかに記載の音声認識方式。

【請求項7】 上記モデルを音韻のモデルおよび音韻境界のモデルとしたことを特徴とする請求項1ないし4いずれかに記載の音声認識方式。

【請求項8】 上記音韻のモデルをセミ連続分布のモデルとし、音韻のモデルの分岐係数だけを適応化することを特徴とする請求項5または7いずれかに記載の音声認識方式。

【請求項9】 上記音韻境界のモデルをセミ連続分布のモデルとし、音韻境界のモデルの分岐係数だけを適応化することを特徴とする請求項6または7いずれかに記載の音声認識方式。

【請求項10】 上記音韻境界のモデルの入力音声への適応化において、入力音声の途中のビーム内の仮説の尤度の大きい方からN個の仮説を選択する際に、モデル間の遷移を有する部分モデル系列の仮説を選択することを特徴とする請求項6記載の音声認識方式。

【請求項11】 ビーム探索の探索の幅をフレームごとに適応化係数を用いて変化させることを特徴とする請求項2ないし4いずれかに記載の音声認識方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、音声認識の精度改善と、探索量の削減に関する。

【0002】

【従来の技術】従来、音声認識の探索量の削減の手法として、ビーム探索が行われている。日本音響学会 平成元年度春季研究発表会 講演論文集1（平成元年3

月）、5～6頁「DPビームサーチの閾値を入力音声の途中で変更する方法の検討」には、フレーム同期型のDPマッチングにおいて、ビーム探索の閾値を入力音声の途中で変化させることで、探索量を減少させる方法が提案されている。また、特開平6-282295号公報には、観測可能な特徴量を入力とする制御関数を用いてビーム探索の探索範囲を適応的に変化させることで、探索量を減少させる方法が開示されている。ここで、ビーム探索の閾値の制御関数には、ニューラルネットおよび重回帰分析を用いている。

【0003】一方、日本音響学会 平成8年度秋季研究発表会 講演論文集1（平成8年9月）、117～118頁「音声認識のためのN-bestに基づく話者適応化」には、教師なし話者適応化（即時適応化）の方法として、Nベストビタビ認識の結果から推定されたNベストのモデル系列に従って、音韻HMMを連結し、その連結した音韻HMMの入力音声に対する尤度が最大となるように、音韻HMMのパラメータを推定して、認識対象話者に適応化する方法が提案されている。この方法での適応化は、次のステップからなる。

(1)適応化前の音韻HMMを用いて入力音声のNベスト認識を行い、入力音声に対するN個のモデル系列を求める。

(2)各モデル系列ごとに、音韻HMMのパラメータをその話者に適応化する。

(3)適応化後に、最大尤度を示したモデル系列を選択する。

(4)その適応化された音韻HMMのパラメータを用いて現在のHMMを更新する。

上記、ステップ(2)～(4)を繰り返す。従って、上記方法は、入力音声の途中でモデルを変更することができない。

【0004】図14は従来のビーム探索を用いる音声認識方式の機能ブロック図である。音声区間切出手段1によって、入力音声1001から切り出された音声区間の各フレームについて、分析手段2による音声分析を行い、特徴パラメータの時系列3を得る。そして、モデル記憶手段51からパラメータ5aを、また構文情報格納手段4から入力音声に対応するモデルの系列を規定する構文情報をそれぞれ参照し、特徴パラメータの時系列3に対する最適なモデル系列を認識結果1003として、以下のようにして得る。なお、10は入力音声の第1フレームから途中までのフレームに対応する部分モデル系列の仮説を格納する部分モデル系列格納手段である。

【0005】最初のフレーム番号を1、最後のフレーム番号をTとする。まず、最初に、部分モデル系列の初期値を部分モデル系列格納手段10に格納する。次に入力音声のフレーム番号iを1とおく。モデル演算手段7は部分モデル系列格納手段10から部分モデル系列の仮説(Hとする)をとり出す。つぎに、構文情報格納手段4

の構文情報から、部分モデル系列Hに連結可能なモデル（音韻モデルk、複数通りのときもある）を選択し、音韻モデルkに対応するフレーム番号iの特徴パラメータの尤度 $f(k, i)$ を計算する。さらに、音韻モデルkを連結した1フレーム分成長した部分モデル系列の仮説を作成し、ビーム探索用の中間スタック1004に格納する。1フレーム分成長した部分モデル系列の仮説の累積尤度は成長前の種の部分モデル系列の累積尤度に音韻モデルkの尤度を加えたものである。ビーム探索手段9はフレーム番号iについて、中間スタック1004内の部分モデル系列の仮説の累積尤度を相互に比較し、例えば、累積尤度の最大の仮説の尤度を上限とし、この上限からビーム幅8だけ引いた値を下限として、この範囲の累積尤度を有する部分モデル系列の仮説を部分モデル系列格納手段10に格納する。

【0006】この場合、中間スタックからの仮説の選び方としては、例えば、累積尤度の大きい方からNb個の部分モデル系列を選ぶこともできる。ただし、Nbはビーム内に残す仮説の数の最大の数を表す。以上の処理を入力音声の第1フレームから最後のフレームまで行うことによって、部分モデル系列格納手段10には、入力音声の全フレームに対応するモデル系列の仮説がその累積尤度とともに得られる。その後、後向き探索手段1002は、例えば最適な累積尤度の仮説を選ぶことによって、認識結果1003を得る。

【0007】

【発明が解決しようとする課題】入力音声の途中でビーム探索の閾値を変更する従来のビーム探索は、探索量を削減することができるが、認識に用いるモデルのパラメータは一定であり、このようなパラメータが一定のモデルで認識を行うため、認識精度の向上は得られない。また、従来の教師なし適応化は、一定のパラメータのモデルでNベスト認識を行ってN個のモデル系列を求めた後に、認識結果からモデルのパラメータの入力音声への適応化を行う。このため、精度のより高い認識結果を得るためには、適応化された音響モデルにより、再度の認識処理が必要であるという問題があった。

【0008】この発明が解決しようとする課題は、ビーム探索を用いる音声認識において、入力音声の途中で音韻モデルおよび音韻境界のモデルを含むモデルを入力音声に適応化することで、認識の精度を向上させることである。また、入力音声の途中でモデルの適応化を行うとともに、入力音声の途中で得られるモデルの精度改善の結果として、入力音声の途中でビーム探索の幅を絞ることで、探索量を削減することである。

【0009】

【課題を解決するための手段】この発明に係る音声認識方式は、入力音声の各フレームで入力音声にモデルを適応化し、この適応化されたモデルの系列からなるモデル系列に対する入力音声の尤度を探索処理により求め、こ

の尤度に基づき音声認識をするようにしたものである。

【0010】上記探索処理としてはビーム探索が用いられる。

【0011】また、入力音声の途中のビーム内の仮説の尤度の大きい方からN個の仮説を用いて、上記モデルを入力音声へ適応化するようにする。

【0012】また、入力音声の途中のビーム内の仮説の尤度の大きい方からN個の仮説を用いるとともに、尤度に応じた重みによる加重を用いて上記モデルを入力音声へ適応化するようにする。

【0013】また、上記モデルとしては音韻のモデルあるいは音韻境界のモデルとする。

【0014】また、上記音韻のモデルあるいは音韻境界のモデルをセミ連続分布のモデルとし、音韻のモデルの分岐係数だけを適応化する。

【0015】また、上記音韻境界のモデルの入力音声への適応化において、入力音声の途中のビーム内の仮説の尤度の大きい方からN個の仮説を選択する際に、モデル間の遷移を有する部分モデル系列の仮説を選択するようにする。

【0016】また、ビーム探索の探索の幅をフレームごとに適応化係数を用いて変化させるようにする。

【0017】

【発明の実施の形態】

実施の形態1. この実施形態は、モデルとして混合連続分布の音韻モデルを用いる場合を示す。図1は、この実施形態における音声認識方式の機能ブロック図である。入力音声信号1001は音声切出手段1により、例えば10msの一定の分析周期で、例えば25.6msの信号区間であるフレームに分割される。分析手段2は、これをフレームごとに特徴パラメータ3に変換する。フレーム番号tの特徴パラメータを X_t と記す。図2はこれ以降の動作を示すフローチャートである。ステップ21ではモデルの初期化を行う。すなわち適応化前のモデルである初期モデルを初期モデル記憶手段5からモデル記憶手段6にコピーする。また、フレームの番号tを1に設定する。次に、 $t=1$ 番から最終の $t=T$ 番のフレームについて、フレーム番号tを1ずつ増加しながら、フレームごとに以下の処理を行う。

【0018】構文情報格納手段4に格納された構文情報は、部分モデル系列のあとに接続可能なモデルを決めるための情報であり、状態をあらわすノードと、遷移をあらわす枝とから表わされる。これは例えば図3に示すようなグラフとして表現される。またこの構文情報は、構文情報格納手段4内においては図4に示すような表として格納されている。すなわち、ある部分モデル系列の現在の構文的な状態をあらわす番号から、次に接続可能なモデルと、そのモデルを選択したときに拡張された部分モデル系列の次の状態番号が、表として与えられている。図3に対応するグラフの構文状態の遷移表は図4の

ようになる。

【0019】モデル演算手段7は、1フレーム分の入力を行い（ステップ22）、フレームごとに、構文情報に従ったモデルのパラメータを適用し（ステップ23）、入力される特徴パラメータ3の尤度を計算する（ステップ24）。モデルのパラメータは、音韻モデル k について、 M 混合のガウス分布の平均、分散、分岐係数 $\mu_m(k)$ 、 $\Sigma_m(k)$ 、 $\lambda_m(k)$ ($m=1, 2, \dots, M$) からなる。現在の構文状態が p のとき、構文情報から自己ループを含めて後続の遷移可能なすべての枝を検知し、このすべての枝について、その枝のモデルと遷移先の構文状態の組み合わせ $\langle k, q \rangle \in \{\langle k1, q1 \rangle, \langle k2, q2 \rangle, \dots, \langle kn, qn \rangle\}$ に対するモデル k の特徴パラメータ x_t の尤度 $f(t, k)$ を、混合分布の各分布の尤度 $N(x_t, \mu_m(k), \Sigma_m(k))$ の加重和として次式で計算する。

【0020】

【数1】

$$f(t, k) = \sum_{m=1}^M \lambda_m(k) N(x_t, \mu_m(k), \Sigma_m(k))$$

【0021】ステップ25では、次のようにして、1フレーム前の部分モデル系列を1フレーム分拡張し、新しい部分モデル系列を生成する。種となる1フレーム前の部分モデル系列が $S1, S2, \dots$ のとき、部分モデル系列を一つ選択し、 S とする。 S は構文状態 $\delta(S)$ と、累積尤度 $\alpha(S)$ と、最終モデル $k(S)$ とを情報として保持している。 S の構文状態が p のとき、つぎの演算を行い、構文状態、選択されるモデルの組み合わせに応じて、新しい部分モデル系列の仮説 $U1, U2, \dots$ を生成する。例えば、選択されるモデルが k で、次の構文状態が q のとき、これに対応して生成される新しい部分モデル系列を U とすると、 U の構文状態 $\delta(U)$ は $\delta(U)=q$ 、 U の累積尤度 $\alpha(U)$ は $\alpha(U)=\alpha(S)+f(t, k)$ 、 U の最終モデルは $k(U)=k$ である。

【0022】ビーム探索手段9は、モデル演算手段7で生成された部分モデル系列 $U1, U2, \dots$ について、それらの累積尤度と、制御手段1000より与えられるビーム幅8とで決まる、ビーム幅範囲の中に入らない仮説を破棄することで、ビーム幅の中に入る仮説だけを残し、 i モデル系列として出力し、部分モデル系列格納手段10に格納する（ステップ26）。ビーム幅8に基づくビーム幅範囲の設定は、 $U1, U2, \dots$ の累積尤度の中の最大値 α_{max} を上限として、 α_{max} からビーム幅8を減じた値を下限 α_{min} とすることで行う。枝刈りは、 $U1, U2, \dots$ の中から、その累積尤度 $\alpha(U1), \alpha(U2), \dots$ が、 α_{min} から α_{max} の間にある仮説を残し、それ以外を破棄することで行う。

【0023】次に部分モデル系列選択手段11は、適応化手段13における適応化に用いる情報としての部分モデル系列を選択する（ステップ27）。例えば、部分モデル系列格納手段10の中の部分モデル系列で、累積尤度の大きい部分モデル系列から順番に探索し、異なるモデルを選択した部分モデル系列を最大で N 個選択する。

【0024】次に適応化手段13は、部分モデル系列選択手段11が現在のフレームで選択した部分モデル系列 $U1, U2, \dots$ （最大で N 個）の、選択されたモデル $k \in k(U1), k(U2), \dots$ （最大で N 個）について、適応化係数12に従って、パラメータの適応化を行う（ステップ28）。この実施形態においては、モデルのパラメータは、音韻モデル k について、 M 混合のガウス分布の平均、分散、分岐係数 $\mu_m(k)$ 、 $\Sigma_m(k)$ 、 $\lambda_m(k)$ ($m=1, 2, \dots, M$) からなる。適応化の対象は、 M 混合の各分布（正規密度関数）の尤度に対する分岐係数 $\lambda_m(k)$ と、 M 混合の各分布の平均 $\mu_m(k)$ であり、従って、補正前のパラメータ14は、モデル k について $\lambda_m(k)$ と、 $\mu_m(k)$ であり、その適応化は、次式で行う。

【0025】

【数2】

$$\mu_m(k) \leftarrow \mu_m(k) * (1 - w) +$$

$$\frac{x_t \gamma_m}{\sum_{m=1}^M \gamma_m} * w$$

$$(m = 1, 2, \dots, M)$$

$$\lambda_m(k) \leftarrow \lambda_m(k) * (1 - w) + \gamma_m * w$$

$$(m = 1, 2, \dots, M)$$

ここで、

$$\gamma_m = \frac{\lambda_m(k) N(x_t, \mu_m(k), \Sigma_m(k))}{\sum_{m=1}^M \lambda_m(k) N(x_t, \mu_m(k), \Sigma_m(k))}$$

【0026】なお、 w は適応化係数12で $0 \leq w < 1$ 。

分散の適応化は理論上は次式で可能であるが、適応化の

対象となるパラメータ数を削減するため、この実施形態*

$$\Sigma_m(k) \leftarrow \Sigma_m(k) * (1 - w) + \frac{(x_t - \mu_m(k))(x_t - \mu_m(k))^t \gamma_m}{\sum_{m=1}^M \gamma_m} * w$$

($m = 1, 2, \dots, M$)

【0028】全ての選択されたモデルについて、上記の適応化が終了した後、適応化手段13は適応化の結果得られたパラメータを補正後パラメータ15としてモデル記憶手段6のパラメータを補正後のパラメータ15に置き換える(ステップ29、30)。そして、後向き探索手段1002による後向き探索を行い、認識結果1003を得る(ステップ31)。なお、制御部1000はモデル記憶手段6の初期化から、入力フレームに同期したモデル演算手段7の処理、ビーム探索手段9、適応化手段13の各処理の制御を行う。

【0029】以上のように、 t 番目のフレームでの入力フレームの尤度計算に用いるモデルのパラメータは、一つ前のフレームで適応化処理により補正されたパラメータを用いている。これにより、次第に適応化が進んでいく。すなわち、認識結果が出たあとではなく、認識処理中に適応化が進められるものである。また、構文情報を備えるビーム探索の過程の中で、構文情報で規定される部分モデル系列から、尤度の高い部分モデル系列のモデルを適応化の対象のモデルとして選択しているため、いわば過去の履歴で補正されたフレームごとの認識結果によるモデルの適応化が実現されることになっている。このため、従来のビーム探索のビーム幅の制御による、探索量の減少の効果とともに、従来は得られなかった認識精度の改善の効果が期待できる。また、部分モデル系列選択手段11において、累積尤度の大きい部分モデル系列から順番に探索し、異なるモデルを選択した部分モデル系列を最大で N 個選択するようにしたので、安定した適応化が行える。

【0030】実施形態2. 次に、モデルとしてセミ連続分布の音韻モデルを用いる実施形態を示す。この場合のブロック図は図1と同じであり、フローチャートは図2と同じである。モデルが異なるため、モデル演算と適応化部の動作が異なるが、それ以外は同じであり、説明を省略する。

【0031】モデル演算手段7は、フレームごとに、構文情報4に従ったモデルのパラメータを適用し、入力の特徴パラメータ3の尤度を計算する。この実施形態のモデルのパラメータは、すべての音韻について共通の M 個

*では行わない。

【0027】

【数3】

10 のコードブックのガウス分布の平均、分散 μ_m, Σ_m ($m=1, 2, \dots, M$)と、音韻モデル k についての分岐係数 $\lambda_m(k)$ からなる。現在の構文状態が p のとき、構文情報から自己ループを含めて後続の遷移可能なすべての枝を検知し、このすべての枝について、その枝のモデルと遷移先の構文状態の組み合わせ $\langle k, q \rangle \in \{ \langle k1, q1 \rangle, \langle k2, q2 \rangle, \dots, \langle kn, qn \rangle \}$ に対するモデル k の特徴パラメータ x_t の尤度 $f(t, k)$ を、混合分布の各分布の尤度 $N(x_t, \mu_m, \Sigma_m)$ の加重和として次式で計算する。

【0032】

20 【数4】

$$f(t, k) = \sum_{m=1}^M \lambda_m(k) N(x_t, \mu_m, \Sigma_m)$$

【0033】種となる一フレーム前の部分モデル系列が $S1, S2, \dots$ のとき、部分モデル系列の一つを選択し、 S とする。 S は構文状態 $\delta(S)$ と、累積尤度 $\alpha(S)$ と、最終モデル $k(S)$ とを情報として保持している。 S の構文状態が p のとき、つぎの演算を行い、構文状態、選択されるモデルの組み合わせに応じて、新しい部分モデル系列の仮説 $U1, U2, \dots$ を生成する。例えば、選択されるモデルが k で、次の構文状態が q のとき、これに対応して生成される新しい部分モデル系列を U とすると、 U の構文状態 $\delta(U)$ は $\delta(U)=q$ 、 U の累積尤度 $\alpha(U)$ は $\alpha(U)=\alpha(S)+f(t, k)$ 、 U の最終モデルは $k(U)=k$ である。

【0034】適応化手段13は、部分モデル系列選択手段11が現在のフレームで選択した部分モデル系列 $U1, U2, \dots$ (最大で N 個)の、選択されたモデル $k \in k(U1), k(U2), \dots$ (最大で N 個)について、適応化係数12に従ってパラメータの適応化を行う。この実施形態のモデル k のパラメータは、すべての音韻について共通の M 個のコードブック(正規密度関数、平均、分散 Σ_m, λ_m ($m=1, 2, \dots, M$))である。適応化対象は音韻モデル k についての分岐係数 $\mu_m(k)$ である。従って、補正前のパラメータ14は、モデル k について $\lambda_m(k)$ であり、その適応化は次式で行う。

【0035】

【数5】

$$\lambda_m(k) \leftarrow \lambda_m(k) * (1 - w) + \gamma_m * w$$

$$(m = 1, 2, \dots, M)$$

ここで、

$$\gamma_m = \frac{\lambda_m N(x_t, \mu_m, \Sigma_m)}{\sum_{m=1}^M \lambda_m N(x_t, \mu_m, \Sigma_m)}$$

【0036】なお、 $N(x_t, \mu_m, \Sigma_m)$ が第 m 番目のコードブックの尤度（正規密度関数の値）である。 $\lambda_m=0$ なる分岐係数は、適応化しても $\lambda_m=0$ のままである。この実施形態では、したがって、 $\lambda_m=0$ なる係数についての適応化のための演算を省略することで、精度に影響を与えずに、演算量を削減することができる。すべてのモデルについて、上記の適応化が終了した後、適応化手段13は、適応化の結果得られたパラメータを補正後パラメータ15としてモデル記憶手段6のパラメータを補正後のパラメータ15に置き換える。

【0037】以上のように、実施形態1と同様、 t 番目のフレームでの入力フレームの尤度計算に用いるモデルのパラメータは、一つ前のフレームで適応化処理により補正されたパラメータを用いている。これにより、次第に適応化が進んでいく。すなわち、認識結果がでたあとではなく、認識処理中に適応化が進められるものである。また、構文情報を備えるビーム探索の過程の中で、構文情報で規定される部分モデル系列から、尤度の高い部分モデル系列のモデルを、適応化の対象のモデルとして選択しているため、いわば過去の履歴で補正されたフレームごとの認識結果によるモデルの適応化が実現されることになっている。このため、従来のビーム探索のビーム幅の制御による、探索量の減少の効果とともに、従来は得られなかった認識精度の改善の効果が期待できる。この実施形態では、セミ連続分布を用いたため、分岐係数の適応化だけで精度が改善される。計算、適応化が容易である。

【0038】実施形態3. 次に、音韻のモデルについて、フレームごとに適応化を行うもので、モデル系列の尤度に応じた適応化係数による適応化をする実施形態を示す。

$$\lambda_m(k) \leftarrow \lambda_m(k) * (1 - w(k)) + \gamma_m * w(k)$$

$$(m = 1, 2, \dots, M)$$

ここで、

$$\gamma_m = \frac{\lambda_m N(x_t, \mu_m, \Sigma_m)}{\sum_{m=1}^M \lambda_m N(x_t, \mu_m, \Sigma_m)}$$

【0044】なお、 $N(x_t, \mu_m, \Sigma_m)$ が第 m 番目のコードブックの尤度（正規密度関数の値）である。 $\lambda_m=0$ なる分岐係数は、適応化しても $\lambda_m=0$ のままである。

*【0039】この場合のブロック図は図1と同じであり、フローチャートは図2と同じである。音韻のモデルは、実施形態2と同様のセミ連続分布モデルである。この実施形態では音韻モデルとしてセミ連続分布モデルについて説明したが、混合連続分布モデルでも、同様な効果が期待できる。適応化手段13の動作が異なる以外は実施形態2と同様であり、説明を省略する。

【0040】適応化手段13は、部分モデル系列選択手段11が現在のフレームで選択した部分モデル系列 U_1, U_2, \dots （最大で N 個）の、選択されたモデル $k \in k(U_1), k(U_2), \dots$ （最大で N 個）について、適応化係数12に従って、選択された部分系列の尤度に応じて、パラメータの適応化を行う。モデル k について、適応化係数 $w(k)$ の適応化を行う。ここで、モデル k の適応化係数 $w(k)$ は、

【0041】

【数6】

$$w(k) = w \times \frac{\alpha(U(k))}{\sum_{k \in \{\text{対象モデル}\}} \alpha(U(k))}$$

【0042】式中、 $U(k)$ は選択されたモデル k を選択するにあたって用いられた部分モデル系列である。この実施形態のモデル k のパラメータは、すべての音韻について共通の M 個のコードブック（正規密度関数、平均、分散 μ_m, Σ_m ($m=1, 2, \dots, M$))である。適応化対象は、音韻モデル k についての分岐係数 $\lambda_m(k)$ である。従って、補正前のパラメータ14は、モデル k について $\lambda_m(k)$ であり、その適応化は、次式で行う。

【0043】

【数7】

この実施形態では、したがって、 $\lambda_m=0$ なる係数についての適応化のための演算を省略することで、精度に影響を与えずに、演算量を削減することができる。すべて

のモデルについて、上記の適応化が終了した後、適応化手段13は、適応化の結果得られたパラメータを補正後パラメータ15としてモデル記憶手段6のパラメータを補正後のパラメータ15に置き換える。

【0045】以上のように、実施形態1と同様、t番目のフレームでの入力フレームの尤度計算に用いるモデルのパラメータは、一つ前のフレームで適応化処理により補正されたパラメータを用いている。これは、構文情報を備えるビーム探索の過程の中で、構文情報で規定される部分モデル系列から、尤度の高い部分モデル系列のモデルを、適応化の対象のモデルとして選択しているため、いわば過去の履歴で補正されたフレームごとの認識結果によるモデルの適応化が実現されることになっている。このため、従来のビーム探索のビーム幅の制御による、探索量の減少の効果とともに、従来は得られなかった認識精度の改善の効果が期待できる。この実施形態ではセミ連続分布を用いたため、分岐係数の適応化だけで*

$$Pr(B_t|境界) = \sum_{m=1}^{MB} P_m N(B_t, \mu_m, \Sigma_m)$$

$$Pr(B_t|非境界) = \sum_{m=1}^{MB} Q_m N(B_t, \mu_m, \Sigma_m)$$

ただし、

$$\sum_{m=1}^{MB} P_m = 1$$

$$\sum_{m=1}^{MB} Q_m = 1$$

かつ

$$0 \leq P_m \leq 1 (m = 1, 2, \dots, MB)$$

$$0 \leq Q_m \leq 1 (m = 1, 2, \dots, MB)$$

【0048】この実施形態での部分モデル系列選択手段11は、部分モデル系列格納手段10の部分モデル系列の中から、音韻境界の遷移が起こった部分モデル系列（即ち、自己ループに対応しないもの）を尤度の大きい方から、最大でN個選択する。これにより、特別な計算をすることなく選択が行える。また、この実施形態での適応化手段13は、部分モデル系列選択手段11が現在のフレームで選択した部分モデル系列U1, U2, ...（最大でN個）の、選択されたモデル $k \in k(U1), k(U2),$

*精度が改善される。計算、適応化が容易である。また、部分系列の尤度を考慮するため、誤った方向の適応化を防止することが期待できる。

【0046】実施形態4. 次に音韻境界のモデルについて、フレームごとに適応化を行うものを示す。音韻境界のモデルは、音韻間の遷移に対応したモデル間の遷移を制御するためのモデルであり、次の尤度比が1より大きいときに音韻間の遷移が可能である。

<尤度比> = <音韻境界である第1の確率密度> / <音韻境界でない第2の確率密度>

この実施形態では、第1の確率密度($Pr(B_t | 境界)$)および第2の確率密度($Pr(B_t | 非境界)$)は、コードブックの確率密度関数の次の多項式で与えられる。但し、 B_t はt番目及びその前後のフレームから作成した特徴量である。

【0047】

【数8】

...（最大でN個）について、適応化係数12に従って、パラメータの適応化を行う。この実施形態の音韻境界モデルkのパラメータは、コードブックの尤度に対する分子多項式係数 $P_m(k)$ であり、従って、補正前のパラメータ14はモデルkについて $P_m(k)$ であり、その適応化は次式で行う。

【0049】

【数9】

$$P_m(k) \leftarrow P_m(k) * (1 - w) + \gamma_m * w$$

$$(m = 1, 2, \dots, MB)$$

ただし、

$$\gamma_m = \frac{P_m N(B_t, \mu_m, \Sigma_m)}{\sum_{m=1}^{MB} P_m N(B_t, \mu_m, \Sigma_m)}$$

【0050】なお、MBは音韻境界モデル用のコードブック（正規密度関数）の数、 $N(B_t, \mu_m, \Sigma_m)$ は正規密度関数、 μ_m, Σ_m はそれぞれ正規密度関数の平均および分散である。wは適応化係数である。Pm=0なる多項式係数は、適応化してもPm=0のままである。この実施形態では、したがって、Pm=0なる係数についての適応化のための演算を省略することで、精度に影響を与えずに演算量を削減することができる。すべてのモデルについて、上記の適応化が終了した後、適応化手段13は、適応化の結果得られたパラメータを補正後パラメータ15としてモデル記憶手段6のパラメータを補正後のパラメータ15に置き換える。

【0051】実施形態5. 次に、フレームごとのモデルの適応化処理とともに、ビーム探索の幅を、フレームに同期して、漸減させる例を示す。図5にビーム探索の幅の変化を模式的に示す。フレームごとのモデルの適応化処理によって、尤度が高くなることが期待され、ビーム内における正解の仮説の順位が向上する。このため、ビーム幅をフレームごとに漸減させることで、探索量が削減される。ビーム幅8の更新は次式で行う。但し、 θ はビーム幅である。

$$\theta \leftarrow \theta * (1 - w) + \langle \text{ビーム幅推定値} \rangle * w$$

【0052】ここで、 $\langle \text{ビーム幅推定値} \rangle$ は、数多くの例について認識実験を行い、最終入力フレームにおいて正解の部分モデル系列の尤度と、そのときの尤度が最大の部分モデル系列の尤度との差として求めた。ビーム幅の初期値は、 $\langle \text{ビーム幅推定値} \rangle$ に比べ、大きな値に設定する。上の式でwはビーム幅をフレームごとに更新するときの度合いを決める適応化係数である。

【0053】適応化係数をどのように設定するのが妥当かを実験的に決めるため評価実験を行った。音節の3連鎖の制約を構文情報とする。出力の仮説はグラフ構造になっている。グラフ構造の複雑さの減少の程度でフレームごとの適応化の効果を調べた。図6は、音韻モデル（セミ連続分布モデル）の1フレームごとの適応化を行う実施形態2の適応化のため選択する仮説数Nと適応化係数wとの組み合わせ条件について、

- (1) 正解のモデル系列の尤度と最大の尤度を示したモデル系列の尤度との差(Δ)
- (2) 出力グラフのノード数
- (3) 出力グラフのエッジ数（枝の数）

に基づいて作成した実験結果を示す。

【0054】それぞれの数値は、(1)については、フレームごとの適応化なしの場合を0として、それに対する Δ の増加分を、また、(2)と(3)については、フレームごとの適応化なしの場合を1として、それに対する比を、様々な不特定話者の入力音声を確認したときについて平均した数値を示す。なお、評価に用いた入力音声は、次の20フレーズである。

【0055】（話者）：（フレーズの音韻記述）

ec10009 : kaizjoowa dociradesuka

ec10009 : kikaisiNkookaikaNnara tookjootawaano mae desu

ec10009 : tookjootawaano maedesuka

et11003 : tookjootawaano maedesu

et11003 : tookjootawaano mae

fuj0003 : koNdonu hujujasumini

fuj0003 : cukubani cuite osiete kudasai

fuj0003 : cukubawa

fuj0003 : zjeeaarude kuru baaiwa

kdd1005 : koNdo

30 kdd1005 : oNseekerNkjuukaiga aruNde soreo kikini ik itaiNdesukeredo

mac0003 : kikaisiNkookaikaNdesu

mat1003 : koNdo oNseekerNkjuukaiga aruNde

mat1003 : tookjootawaadesu

mit0003 : kanazawano rjokooaNnaisjodesjooka

mit0003 : sinainiwa cjuuooni keNrokueNga arimasu

nec1011 : kaNkoopuraNzukurio otasukesimasu

nec1011 : dokoka mite mitai tokorowa arimasuka

nec1011 : rakuhoikuhoomeNto

40 【0056】また図6の結果をグラフにして表現したものを図7、図8、図9に示す。それぞれX軸を適応化係数w、Y軸を適応化のため選択する仮説数Nとしたものであり、Z軸として図7は上記(1)の Δ 、図8は上記(2)のノード数、図9は上記(3)のエッジ数をとったものである。なお、XY平面上にはZ軸の等高線を示している。図6～図9から、w=0.005かつN=1～50、また、w=0.01かつN=1～50、さらに、w=0.02かつN=1～200、w=0.05かつN=50～100にすれば、 Δ が減少かつノード数とエッジ数が減少することがわかる。 Δ の減少
50 は音声認識の精度の向上を示し、またノード数とエッジ

数の減少は、音声認識の精度の向上によって、正解以外のモデル系列の生成が抑制されたことを示していると考えられる。

【0057】図10は、音韻境界のモデル（セミ連続分布モデル）の1フレームごとの適応化を行う実施形態4の適応化のため選択する仮説数 N と適応化係数 w との組み合わせ条件について、

(1)正解のモデル系列の尤度と最大の尤度を示したモデル系列の尤度との差(Δ)

(2)出力グラフのノード数

(3)出力グラフのエッジ数（枝の数）

に基づいて作成した実験結果を示す。

【0058】それぞれの数値は、(1)については、フレームごとの適応化なしの場合を0として、それに対する Δ の増加分を、また、(2)と(3)については、フレームごとの適応化なしの場合を1として、それに対する比を、様々な不特定話者の入力音声を確認したときについて平均した数値を示す。評価に用いた入力音声は、上記の20フレーズである。また図10の結果をグラフにして表現したものを図11、図12、図13に示す。それぞれX軸を適応化係数 w 、Y軸を適応化のため選択する仮説数 N としたものであり、Z軸として図11は上記(1)の Δ 、図12は上記(2)のノード数、図13は上記(3)のエッジ数をとったものである。なお、XY平面上にはZ軸の等高線を示している。

【0059】図10～図13から、音韻境界モデルの適応化係数 w と適応化する境界の種類数 N の適切な範囲としては、 $w=0.1$ かつ $N=100\sim500$ 、また、 $w=0.2$ かつ $N=100$ 、さらに、 $w=0.3$ かつ $N=50\sim500$ 、 $w=0.4$ かつ $N=50\sim500$ 、 $w=0.5$ かつ $N=1\sim500$ などで、 Δ が減少かつノード数とエッジ数が減少することがわかる。 Δ の減少は、音声認識の精度の向上を示し、また、ノード数とエッジ数の減少は、音声認識の精度の向上によって、正解以外のモデル系列の生成が抑制されたことを示していると考えられる。

【0060】

【発明の効果】以上に説明したように、この発明によれば、入力音声の各フレームで入力音声にモデルを適応化し、この適応化されたモデルの系列からなるモデル系列に対する入力音声の尤度を探索処理により求めるようにしたので、過去の履歴で補正されたフレームごとの認識結果によるモデルの適応化が実現されることになり、探索量の減少の効果とともに認識精度の改善の効果がある。

【0061】また、入力音声の途中のビーム内の仮説の尤度の大きい方から N 個の仮説を用いて、上記モデルを入力音声へ適応化することにより、安定した適応化が行える。

【0062】また、入力音声の途中のビーム内の仮説の尤度の大きい方から N 個の仮説を用いるとともに、尤度

に応じた重みによる加重を用いて上記モデルを入力音声へ適応化することにより、安定した適応化が行える。

【0063】また、上記モデルを音韻のモデルのセミ連続分布のモデルとし、音韻のモデルの分岐係数だけを適応化することにより、計算処理や適応化を容易に行うことができる。

【0064】また、上記モデルを音韻境界のモデルのセミ連続分布のモデルとし、音韻境界のモデルの分岐係数だけを適応化することにより、計算処理や適応化を容易に行うことができる。

【0065】また、音韻境界のモデルの入力音声への適応化において、入力音声の途中のビーム内の仮説の尤度の大きい方から N 個の仮説を選択する際に、モデル間の遷移を有する部分モデル系列の仮説を選択することにより、容易に選択を行える。

【0066】また、ビーム探索の探索の幅をフレームごとに適応化係数を用いて変化させることにより、ビームの幅をモデルの適応の度合いに応じて絞り、探索量を削減することができる。

10 【図面の簡単な説明】

【図1】 この発明の実施形態における音声認識方式の機能ブロック図である。

【図2】 この発明の実施形態における音声認識動作のフローチャートである。

【図3】 この発明の実施形態における構文制御情報の模式図である。

【図4】 この発明の実施形態における構文制御情報の構成の説明図である。

【図5】 この発明の実施形態におけるビーム探索の幅の変化を示す説明図である。

【図6】 この発明の実施形態における評価結果の説明図である。

【図7】 この発明の実施形態における評価結果をグラフ化して示す説明図である。

【図8】 この発明の実施形態における評価結果をグラフ化して示す説明図である。

【図9】 この発明の実施形態における評価結果をグラフ化して示す説明図である。

【図10】 この発明の実施形態における評価結果の説明図である。

【図11】 この発明の実施形態における評価結果をグラフ化して示す説明図である。

【図12】 この発明の実施形態における評価結果をグラフ化して示す説明図である。

【図13】 この発明の実施形態における評価結果をグラフ化して示す説明図である。

【図14】 従来の音声認識方式の機能ブロック図である。

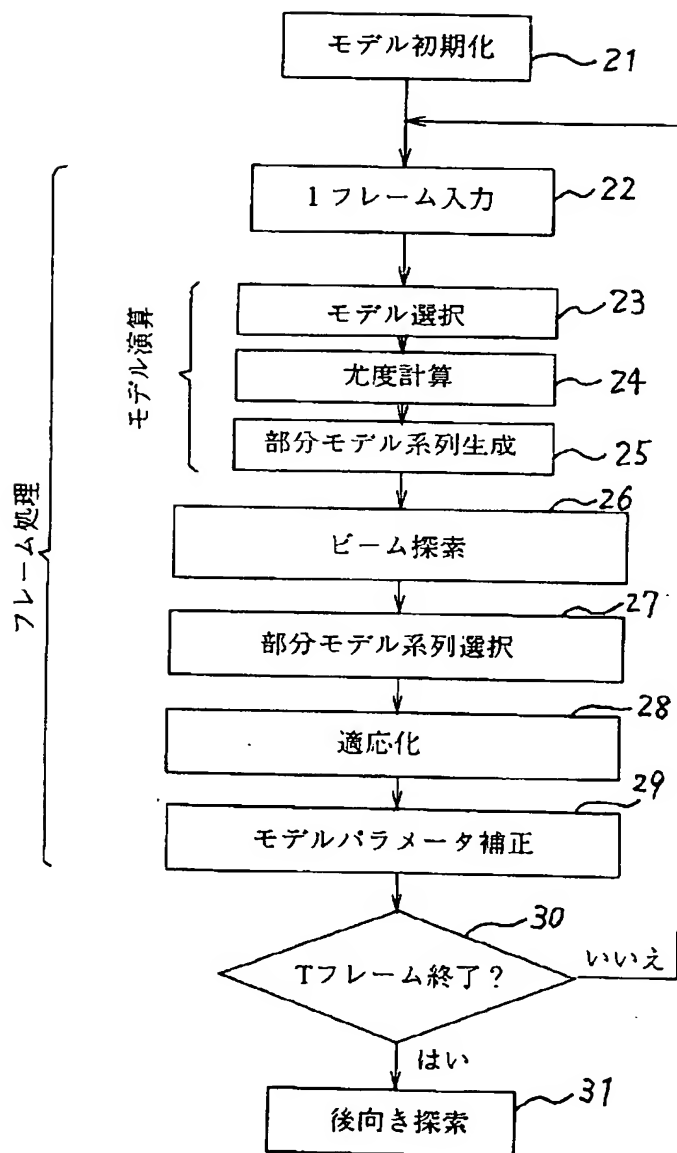
【符号の説明】

1 音声区間切出手段

- 2 分析手段
- 3 特徴パラメータ
- 4 構文情報格納手段
- 5 初期モデル記憶手段
- 6 モデル記憶手段
- 7 モデル演算手段
- 8 ビーム幅
- 9 ビーム探索手段
- 10 部分モデル系列格納手段
- 11 部分モデル系列選択手段

- 12 適応化係数
- 13 適応化手段
- 14 補正前パラメータ
- 15 補正後パラメータ
- 1000 制御手段
- 1001 入力音声
- 1002 後向き探索手段
- 1003 認識結果
- 1004 中間スタック

【図2】

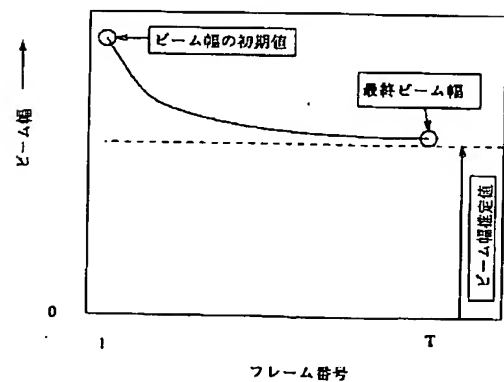


【図4】

初期状態の番号 0	
最終状態の番号 1	

現在の状態番号	選択される音素モデル	次の状態番号
2	t	3
4	u	5
4	i	2
3	a	5
6	r	7
6	t	3
8	e	6
9	r	8
9	s	4
10	a	9
5	#	1
11	r	7
12	u	11
13	j	12
13	w	10
13	u	5
13	Q	2
14	a	13
15	r	14
16	a	15
0	#	16

【図5】

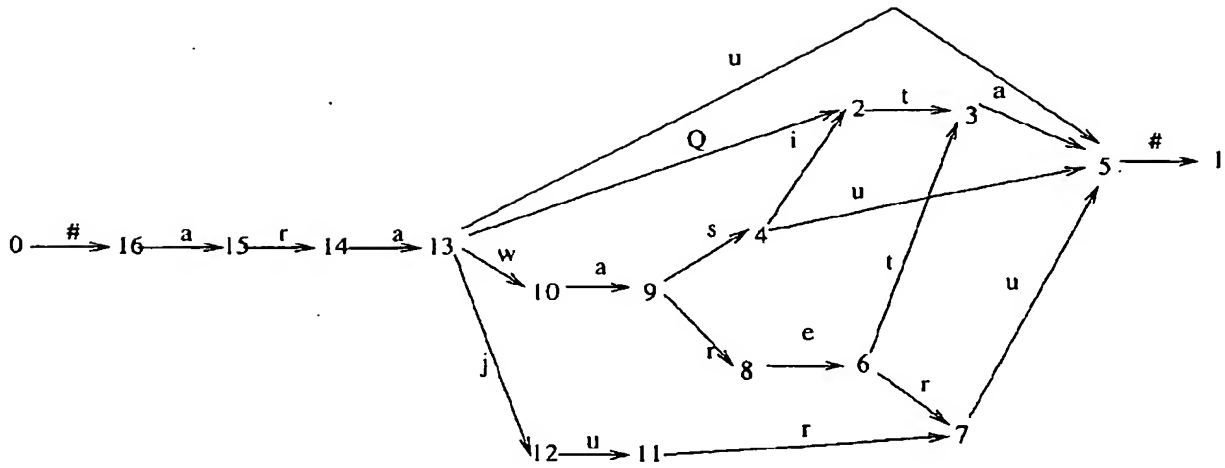


(11)

特開平 1 0 - 2 5 4 4 9 6

【図 1】

【図3】

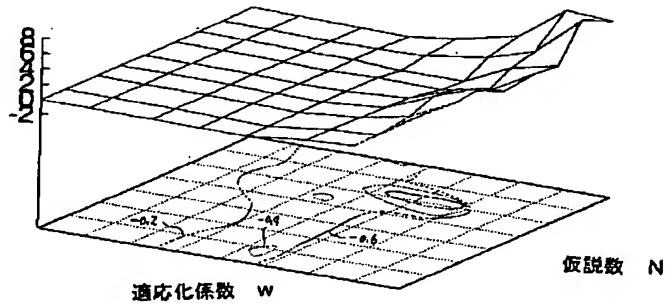


【図6】

W	N	誤差の増分	ノード数の比	エッジ数の比
0.000	1	0.000	1.000	1.000
0.001	2	-0.046	0.982	0.988
0.001	5	-0.046	0.988	0.988
0.001	10	-0.037	0.992	0.989
0.001	20	-0.046	0.991	0.989
0.001	50	-0.036	1.000	0.998
0.001	100	-0.001	1.006	1.004
0.001	200	-0.093	1.058	1.063
0.001	500	-0.119	1.165	1.200
0.002	1	-0.083	0.982	0.978
0.002	2	-0.082	0.982	0.978
0.002	5	-0.069	0.982	0.978
0.002	10	-0.069	0.982	0.978
0.002	20	-0.094	0.985	0.980
0.002	50	-0.109	0.991	0.985
0.002	100	-0.174	1.009	1.007
0.002	200	-0.154	1.094	1.108
0.002	500	-0.181	1.259	1.308
0.005	1	-0.114	0.967	0.957
0.005	2	-0.180	0.967	0.957
0.005	5	-0.151	0.954	0.944
0.005	10	-0.141	0.954	0.946
0.005	20	-0.192	0.959	0.947
0.005	50	-0.329	0.981	0.976
0.005	100	-0.301	1.026	1.020
0.005	200	-0.277	1.171	1.198
0.005	500	-0.309	1.598	1.685
0.010	1	-0.317	0.941	0.926
0.010	2	-0.317	0.941	0.926
0.010	5	-0.285	0.923	0.917
0.010	10	-0.268	0.940	0.923
0.010	20	-0.323	0.945	0.927
0.010	50	-0.479	0.973	0.962
0.010	100	-0.360	1.027	1.021
0.010	200	-0.338	1.205	1.241
0.010	500	-0.446	1.619	1.637
0.020	1	-0.369	0.894	0.881
0.020	2	-0.445	0.894	0.881
0.020	5	-0.351	0.895	0.886
0.020	10	-0.328	0.896	0.886
0.020	20	-0.284	0.898	0.886
0.020	50	-0.249	0.933	0.908
0.020	100	-0.217	1.000	0.988
0.020	200	-0.147	1.191	1.222
0.020	500	0.343	1.673	1.678
0.050	1	0.484	0.784	0.760
0.050	2	0.503	0.787	0.763
0.050	5	0.419	0.790	0.760
0.050	10	0.342	0.798	0.747
0.050	20	0.251	0.805	0.750
0.050	50	-0.462	0.878	0.839
0.050	100	-0.724	0.941	0.904
0.050	200	0.982	1.207	1.239
0.050	500	2.036	1.790	1.808
0.100	1	0.814	0.693	0.632
0.100	2	0.852	0.697	0.636
0.100	5	1.568	0.701	0.639
0.100	10	2.870	0.725	0.666
0.100	20	2.877	0.727	0.661
0.100	50	0.413	0.744	0.690
0.100	100	2.045	0.835	0.775
0.100	200	2.820	1.037	1.035
0.100	500	8.413	1.580	1.604
0.200	1	4.801	0.534	0.467
0.200	2	4.801	0.534	0.459
0.200	5	4.438	0.537	0.460
0.200	10	4.947	0.516	0.453
0.200	20	3.141	0.529	0.449
0.200	50	1.690	0.649	0.476
0.200	100	2.337	0.672	0.600
0.200	200	7.859	0.907	0.890
0.200	500	6.044	1.255	1.467

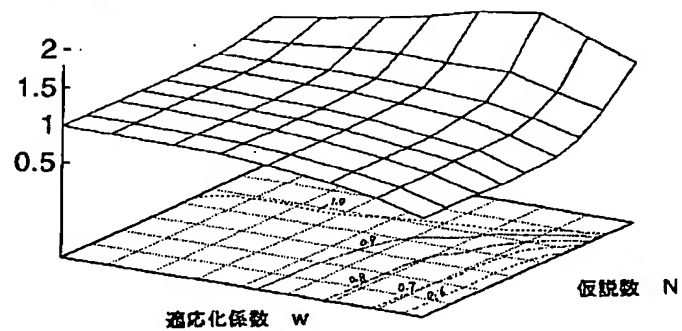
【図7】

log Δ

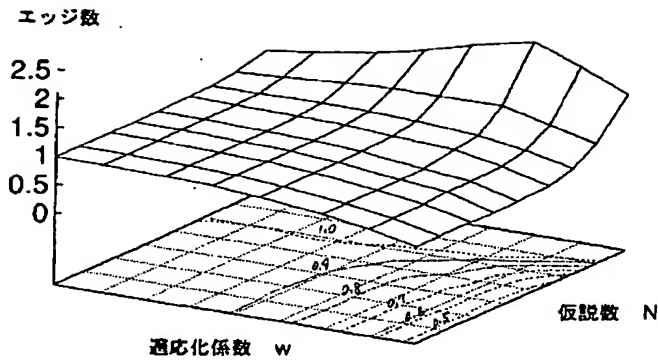


【図8】

ノード数



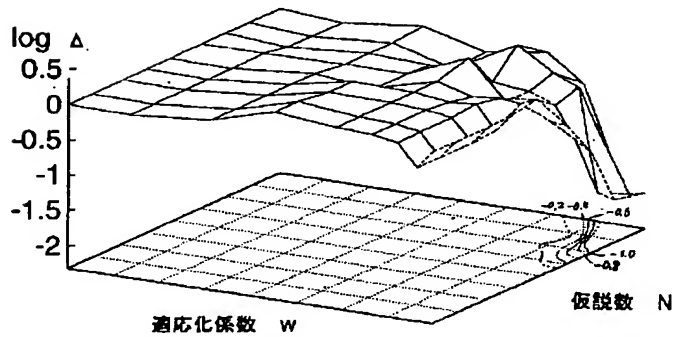
【図9】



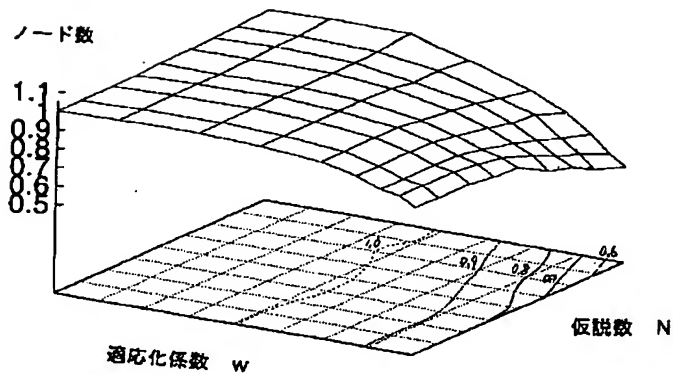
【図10】

w	N	最適化係数 (S) の値	ノード数の比	エッジ数の比
0.000	1	0.000	1.000	1.000
0.010	1	-0.013	1.005	1.004
0.020	2	-0.013	1.005	1.004
0.030	5	-0.013	1.010	1.009
0.050	10	-0.013	1.015	1.017
0.080	20	-0.013	1.012	1.012
0.100	50	0.000	1.010	1.011
0.150	100	-0.003	1.031	1.025
0.200	200	-0.002	1.032	1.028
0.300	500	-0.005	1.051	1.038
0.500	1	0.023	1.011	1.033
0.550	2	0.028	1.011	1.028
0.600	5	0.028	1.014	1.012
0.650	10	0.028	1.020	1.023
0.700	20	0.028	1.014	1.013
0.750	50	0.028	1.017	1.019
0.800	100	0.027	1.016	1.017
0.850	200	0.025	1.020	1.023
0.900	500	0.025	1.044	1.049
0.100	1	0.325	0.995	0.990
0.100	2	0.329	0.994	0.989
0.100	5	0.327	0.995	0.990
0.100	10	0.327	0.996	0.995
0.100	20	0.326	0.998	0.997
0.100	50	0.326	0.991	0.989
0.100	100	-0.004	0.966	0.962
0.100	200	-0.302	0.977	0.970
0.100	500	-0.175	0.956	0.941
0.200	1	0.246	0.965	0.950
0.200	2	0.249	0.968	0.950
0.200	5	0.249	0.967	0.954
0.200	10	0.249	0.967	0.954
0.200	20	0.249	0.949	0.933
0.200	50	0.378	0.930	0.912
0.200	100	-0.112	0.913	0.894
0.200	200	0.234	0.888	0.852
0.200	500	-0.048	0.856	0.825
0.300	1	0.226	0.924	0.899
0.300	2	0.228	0.921	0.896
0.300	5	0.228	0.924	0.901
0.300	10	0.181	0.921	0.883
0.300	20	0.106	0.901	0.871
0.300	50	-0.147	0.870	0.835
0.300	100	-0.164	0.844	0.804
0.300	200	-0.052	0.832	0.799
0.300	500	-0.401	0.770	0.737
0.400	1	0.187	0.875	0.843
0.400	2	0.187	0.875	0.843
0.400	5	0.187	0.879	0.849
0.400	10	0.210	0.873	0.839
0.400	20	0.157	0.851	0.809
0.400	50	-0.070	0.812	0.762
0.400	100	-0.394	0.773	0.722
0.400	200	-0.231	0.732	0.692
0.400	500	-1.333	0.653	0.640
0.500	1	-0.150	0.837	0.832
0.500	2	-0.150	0.860	0.832
0.500	5	-0.150	0.860	0.833
0.500	10	-0.191	0.846	0.812
0.500	20	0.106	0.826	0.789
0.500	50	-0.347	0.764	0.709
0.500	100	-0.742	0.707	0.646
0.500	200	-1.856	0.656	0.611
0.500	500	-1.869	0.592	0.540
0.600	1	0.113	0.810	0.770
0.600	2	0.113	0.808	0.767
0.600	5	0.106	0.810	0.769
0.600	10	0.137	0.794	0.748
0.600	20	0.203	0.780	0.742
0.600	50	-0.360	0.699	0.636
0.600	100	-0.848	0.656	0.604
0.600	200	-1.710	0.598	0.546
0.600	500	-1.814	0.540	0.481

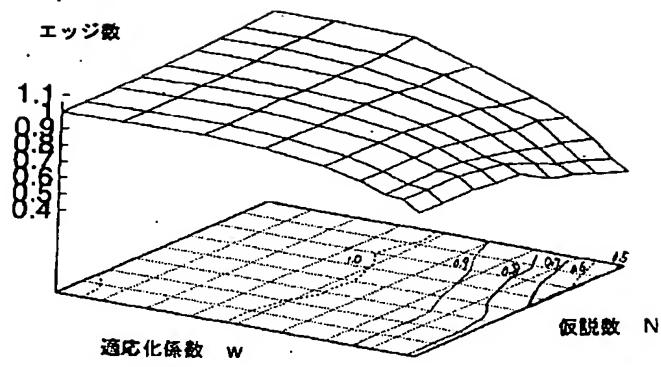
【図11】



【図12】



【図 13】



【図14】

